(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4889618号
(P4889618)

(45)発行日　平成24年3月7日(2012.3.7)　(24)登録日　平成23年12月22日(2011.12.22)

(51)Int.Cl.		F I		
H04L 12/66	(2006.01)	H04L	12/66	B
H04L 12/56	(2006.01)	H04L	12/56	400Z
G06N 3/00	(2006.01)	G06N	3/00	560A

請求項の数 8　(全 24 頁)

(21)出願番号　特願2007-308602 (P2007-308602)
(22)出願日　平成19年11月29日 (2007.11.29)
(65)公開番号　特開2009-135649 (P2009-135649A)
(43)公開日　平成21年6月18日 (2009.6.18)
審査請求日　平成22年9月16日 (2010.9.16)

(73)特許権者　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目７番３号
(74)代理人　100099461
弁理士　溝井　章司
(72)発明者　大野　一広
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内

審査官　安藤　一道

最終頁に続く

(54)【発明の名称】データ処理装置及びデータ処理方法及びプログラム

(57)【特許請求の範囲】

【請求項１】

データを順次入力し、入力したデータの特性を表すデータ特性値を算出するデータ特性値算出部と、

前記データ特性値算出部によりデータ特性値が算出されたデータを順次データ特性値に基づいて分類するデータ分類部と、

前記データ分類部により設けられたデータ類型を計数するとともに、データ類型の増加状況を監視し、データ類型の個数が収束したか否かを判断し、データ類型の個数が収束するまでは、前記データ特性値算出部にデータの入力及びデータ特性値の算出を継続させ、データ類型の個数が収束したと判断した際に、前記データ特性値算出部のデータの入力及びデータ特性値の算出を終了させる収束判定部とを有することを特徴とするデータ処理装置。

【請求項２】

前記収束判定部は、

前記データ分類部により設けられたデータ類型ごとに、データ類型に属するデータ特性値の代表となる代表データ特性値を選択し、選択した各代表データ特性値を出力することを特徴とする請求項１に記載のデータ処理装置。

【請求項３】

前記収束判定部は、

前記データ特性値算出部がデータの入力及びデータ特性値の算出を終了した後に、デー

タ類型ごとに代表データ特性値を選択することを特徴とする請求項２に記載のデータ処理装置。

【請求項４】

前記収束判定部は、

時系列データの異常検知に用いられる学習データとして、各代表データ特性値を出力することを特徴とする請求項２又は３に記載のデータ処理装置。

【請求項５】

前記収束判定部は、

データ類型の増加状況を監視し、データ類型の個数の増加が一定レベル以下に鈍化した場合に、データ類型の個数が収束したと判断することを特徴とする請求項１～４のいずれかに記載のデータ処理装置。

【請求項６】

前記収束判定部は、

データ類型の増加状況を監視し、データ類型の個数が所定の基準値に到達した場合に、データ類型の個数が収束したと判断することを特徴とする請求項１～５のいずれかに記載のデータ処理装置。

【請求項７】

コンピュータが、データを順次入力し、入力したデータの特性を表すデータ特性値を算出するデータ特性値算出ステップと、

前記コンピュータが、前記データ特性値算出ステップによりデータ特性値が算出されたデータを順次データ特性値に基づいて分類するデータ分類ステップと、

前記コンピュータが、前記データ分類ステップにより設けられたデータ類型を計数するとともに、データ類型の増加状況を監視し、データ類型の個数が収束したか否かを判断し、データ類型の個数が収束するまでは、前記データ特性値算出ステップによるデータの入力及びデータ特性値の算出を継続させ、データ類型の個数が収束したと判断した際に、前記データ特性値算出ステップによるデータの入力及びデータ特性値の算出を終了させる収束判定ステップとを有することを特徴とするデータ処理方法。

【請求項８】

データを順次入力し、入力したデータの特性を表すデータ特性値を算出するデータ特性値算出処理と、

前記データ特性値算出処理によりデータ特性値が算出されたデータを順次データ特性値に基づいて分類するデータ分類処理と、

前記データ分類処理により設けられたデータ類型を計数するとともに、データ類型の増加状況を監視し、データ類型の個数が収束したか否かを判断し、データ類型の個数が収束するまでは、前記データ特性値算出処理によるデータの入力及びデータ特性値の算出を継続させ、データ類型の個数が収束したと判断した際に、前記データ特性値算出処理によるデータの入力及びデータ特性値の算出を終了させる収束判定処理とをコンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ネットワークの不正アクセスを監視する技術に関し、特にＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）などの内部ネットワークで発生する通信トラフィックの監視を容易にするために、監視するトラフィックの特徴を学習する学習期間を自動的に判定する技術に関する。

【背景技術】

【０００２】

不正アクセス検出において、収集されたパケットログから生成された時系列データを解析して異常を検知する手法がある。

この手法では、時系列データと学習データとを比較する。学習データとは時系列データ

の変化量を測るための基準となるものである。

【０００３】

学習データを用いた不正アクセス検出技術として、特許文献１及び特許文献２に記載の技術がある。

【０００４】

特許文献１に記載のネットワーク異常検出装置はネットワークの異常検出を行うことを目的とする。

特許文献１に記載のネットワーク異常検出装置は、検出対象たるネットワーク中を一定時間の間に通過するパケットについて、ｋ個（ｋ：自然数）の分類ごとにパケット数をカウントするパケット数カウント部と、カウントしたパケット数をｋ個の分類ごとに正規化し、正規化したパケット数を要素としたｋ次元ベクトルを生成するベクトル生成部とを有する。

さらに、特許文献１のネットワーク異常検出装置は、ｋ次元特徴空間において各次元間の相関関係に基づいて定められた主成分軸を導出する主成分軸導出部と、必要な情報を記憶する記憶部と、生成したｋ次元ベクトルとの距離を導出する主成分軸－ｋ次元ベクトル間距離測定部と、ネットワークの異常の有無を判定する異常判定部とを有する。

このような構成を有することで、特許文献１に記載のネットワーク異常検出装置は、ネットワーク回線の評価を定量的に行うことができるとともに、未知のネットワーク異常をも検出することが可能であるとされる。

【０００５】

また、特許文献２に記載のネットワーク異常検出装置は、アノマリ型侵入検知システムにおいて用いられる学習データを自動的に生成することができる侵入検知システムを提供することを目的とする。

特許文献２では、ネットワーク上を伝送するトラフィックデータを入力し、学習データ作成装置により作成された学習データと前記トラフィックデータから変数を選択する変数選択手段と、該選択された変数をニューラルネットや決定木等の解析アルゴリズムを用いて解析を行い、パターンを生成する処理手段と、該生成されたパターンを用いて前記解析結果を評価する評価手段とを有し、前記変数選択手段および処理手段、評価手段における処理を１回以上行うことにより、侵入の検知に有効なパターンを生成して異常なトラフィックデータを検知する侵入検知システムが記載されている。

特許文献２に記載のネットワーク異常検出装置によれば、アノマリ型ＩＤＳ（Ｉｎｔｒｕｓｉｏｎ　Ｄｅｔｅｃｔｉｏｎ　Ｓｙｓｔｅｍ）に用いる有効なパターンデータの生成に必要な学習データをシグネチャ型ＩＤＳを利用して生成することとしたことから、従来、熟練者等によらなければ入手が困難であった学習データを容易に入手することができるとされる。

また、上記有効なパターンデータをアノマリ型ＩＤＳに適用することにより、シグネチャが登録されていない未知の攻撃や亜種の攻撃を検知できるとされる。

さらに、シグネチャ型ＩＤＳにより生成した学習データを用いて侵入検知に有効なパターンデータを生成し、これをアノマリ型ＩＤＳに適用したことから、より高い検知率を期待できるとされる。

【特許文献１】特開２００４－３１２０６４号公報

【特許文献２】特開２００４－３１２０８３号公報

【発明の開示】

【発明が解決しようとする課題】

【０００６】

特許文献１及び特許文献２に記載のネットワーク異常検知手法では、主に外部ネットワークから内部ネットワークへの脅威を想定している。

それらの脅威はインターネットからの大規模攻撃が主なものである。

内部ネットワークで発生する通信トラフィックは外部ネットワークでの通信トラフィックと比較して小規模な変化である。ネットワークを流れるトラフィックの異常な変動を早

期に捕らえるためには、ネットワーク異常検知システムを検知が可能な状態に早い期間で移行させる必要がある。

それには、システムが正常なネットワークの状態を学習する段階を早期に終了させることが必要である。そのためには、ネットワークの状態の学習期間を終了させるための明確な基準が必要になる。

既存のネットワーク異常検知システムは、ネットワークトラフィックの異常を判断するための比較対象として、先立って正常とシステムに学習させた学習データを用いる。

正常なネットワーク状態の学習にはネットワークが正常に動作していた際のトラフィックデータが一定量必要である。

しかし、学習時の課題として、収集する期間について明確な基準が定義されていない点がある。そのため学習期間の不足や長期化が発生する場合がある。

学習が不足している場合、検知精度が低下する。また学習が長期化した場合、検知に至るまでに余分な時間を要するためシステムの運用に適さない。

そのため学習データは適切に学習をする必要がある。

【０００７】

特許文献１では、前記異常判定手段が、過去のネットワーク状態に基づいて得られた複数のｋ次元ベクトルをクラスタリングによって分類することによって得られた正常領域を用いてネットワークの異常の有無を判定することが示されているが、過去のネットワーク状態をどの程度保有すべきかという基準が存在しない。

また、特許文献２では、ネットワーク上を伝送するトラフィックデータと攻撃種別から攻撃の有無が判定された学習データを作成する学習データ作成装置に関する記載が存在するが、学習を行う期間に対する基準が存在しない。

このため、特許文献１及び特許文献２のいずれにおいても、学習期間が短く、学習データが不足する場合があり、また、逆に、過去の学習期間が長すぎるため、異常検知を開始するタイミングが遅れる場合もある。

このように、特許文献１及び特許文献２では、異常検知のための必要十分な学習期間を決定する基準が存在しないため、十分な学習データが得られず精緻な異常検知を行うことができないという課題、逆に必要以上に学習データの収集に時間を割いた結果、異常検知を効果的なタイミングで行えないという課題がある。

【０００８】

この発明は、このような課題を解決することを主な目的の一つとしており、異常検知に必要十分な学習データが収集された段階で学習データの収集を停止することで、異常検知処理を早期に開始させるとともに、高精度な異常検知を可能とする技術を提供することを主な目的とする。

【課題を解決するための手段】

【０００９】

本発明に係るデータ処理装置は、

データを順次入力し、入力したデータの特性を表すデータ特性値を算出するデータ特性値算出部と、

前記データ特性値算出部によりデータ特性値が算出されたデータを順次データ特性値に基づいて分類するデータ分類部と、

前記データ分類部により設けられたデータ類型を計数するとともに、データ類型の増加状況を監視し、データ類型の個数が収束したか否かを判断し、データ類型の個数が収束するまでは、前記データ特性値算出部にデータの入力及びデータ特性値の算出を継続させ、データ類型の個数が収束したと判断した際に、前記データ特性値算出部のデータの入力及びデータ特性値の算出を終了させる収束判定部とを有することを特徴とする。

【発明の効果】

【００１０】

本発明によれば、データ類型の個数が収束した場合にデータの入力及びデータ特性値の算出を終了させることとしているので、異常検知に必要十分な学習データが収集された段

階で学習データの収集を停止することになり、異常検知処理を早期に開始させることができるとともに、高精度な異常検知を行えるだけの十分な量の学習データを蓄積することができる。

【発明を実施するための最良の形態】

【００１１】

実施の形態１.

以下、本実施の形態では、時系列データから学習データを取得する学習データ取得部７を主に説明する。

先ず、本実施の形態に係る学習データ取得部７が不正アクセスの分析処理においてどのような役割を担っているかを明確にするため、本実施の形態に係る学習データ取得部７が利用される不正アクセス分析システムの概要を説明する。

【００１２】

図１は、本実施の形態に係る学習データ取得部７を含む不正アクセス分析システム１００の構成例を示す。

なお、不正アクセス分析システム１００は、全体として一つのコンピュータで実現されていてもよいし、不正アクセス分析システム１００に含まれる各要素が異なるコンピュータで実現され、各コンピュータがネットワークで接続されて不正アクセス分析システムが実現される形態でもよい。

【００１３】

図１に示す不正アクセス分析システム１００は、例えば図２に示すように、企業等の特定の組織に属するネットワークを監視対象とする。ファイアウォール（Ｆ／Ｗ）、Ｓ－ＮＩＤＳ（Ｓｉｇｎａｔｕｒｅ ｂａｓｅｄ Ｎｅｔｗｏｒｋ ＩＤＳ（Ｉｎｔｒｕｓｉｏｎ Ｄｅｔｅｃｔｉｏｎ Ｓｙｓｔｅｍ））、パケット収集装置からのパケットログ（定点観測データ）を不正アクセス分析システム１００に入力し、リアルタイムに分析を行う。

【００１４】

図１において、情報収集部６は、Ｆ／Ｗ、Ｓ－ＮＩＤＳ、パケット収集装置のパケットログを定期的に収集する。

ログ情報集計部５は、情報収集部６で集められたパケットログから不正アクセスの検知に必要なパケットの情報を集計する。例えば、単位時間当たりの送信元ＩＰアドレス毎パケット数、送信先ポート毎パケット数、或いはパケット長等の集計を行う。

異常検知部４は、ログ情報集計部５により集計されたデータをもとに異常なネットワークトラフィックを検知し早期アラートを出力する。

不正アクセス判定部３は、異常検知部４においてトラフィックの異常状態が検知された場合、不正アクセスが原因であることを判定する機能である。ログ情報集計部５において複数の分析視点での集計を行い、各々に対する異常検知部４の検知の結果を総合的に判断し不正アクセスが原因であることを確定する。また、図示していないセキュリティ情報データベースに格納された既知の脆弱性情報も判定に利用する。例えば、異常検知部４において特定のサービス（ポート）へのパケットの分析結果で異常が検知されており、直近に同サービスの脆弱性が公開されていたのであれば、同脆弱性を悪用した不正アクセスの可能性があると判定できる。

誤検知と判定された場合は、その情報を正常状態して異常検知部４にフィードバックする。

なお、セキュリティ情報データベースとは、例えば、ソフトウェアの最新の脆弱性情報・パッチ情報を管理するデータベースである。

対策部２は、不正アクセス判定部３により不正アクセスが確定された場合、特定ポートへのアクセスの制限、パッチの適用等の指示等、対策の指針を出力する機能である。ネットワーク管理者はこの出力を参考に対策を行う。

ＧＵＩ（Ｇｒａｐｈｉｃａｌ Ｕｓｅｒ Ｉｎｔｅｒｆａｃｅ ）１は、早期アラート、不正アクセスの原因、対策情報等を表示する。

【００１５】
そして、学習データ取得部７は、異常検知部４による異常検知に先立ち、異常検知の対象となる時系列データから所定の時間の間学習データを取得し、取得した学習データを学習データＤＢ（Ｄａｔａ　Ｂａｓｅ）８に格納する。
【００１６】
図３は、本実施の形態に係る学習データ取得部７の学習期間の基準を概念的を示す図である。
【００１７】
本実施の形態に係る学習データ取得部７は、監視先のコンピュータネットワークを流れるトラフィックから、ネットワークに異常が発生していない状態を学習する動作を自動的に完了させる機能を有する。
【００１８】
本実施の形態に係る学習データ取得部７では、パターン数が収束したら学習を終了する。
詳細は後述するが、本実施の形態に係る学習データ取得部７は、時系列データに含まれるデータ部分を複数個のデータ類型に分類する。このデータ類型をパターンという。そして、時系列データに出現するパターンの数が収束した際に学習期間を終了する。
この点、従来の手法では、学習期間をいつ終了するかについての明確な基準が存在しなかったので、十分な数のパターンが出現する前に学習期間を終了してしまい、この結果、学習データが不足し、異常検知における検知精度が低くなる場合があった（図３の左側の従来手法）。
また、逆に、殆どのパターンが出現してしまいパターン数が増えないにもかかわらず学習期間を継続した結果、異常検知を開始するタイミングが遅れ、不正アクセスによりシステムの運用に支障をきたす場合があった（図３の右側の従来手法）。
【００１９】
図４は、本実施の形態に係る学習データ取得部７（データ処理装置）の構成例を示す。
【００２０】
データ入力・処理部７２０は、入力データ７１０を単位時間ごとに集計した数を記憶し、主成分得点計算部７３０に主成分得点計算の対象となるデータを出力する。この入力データ７１０は、学習対象となる時系列データである。なお、以下、入力データ７１０を時系列データ又は学習対象データともいう。
【００２１】
主成分得点計算部７３０は、上記データ入力・処理部７２０で集計された時系列データを順次入力し、入力した時系列データの特性を表すデータ特性値を算出する。以下では、時系列データの主成分得点を計算する例について説明する。主成分得点計算部７３０は、データ特性値算出部の例である。
【００２２】
データ正規化部７４０は、上記主成分得点計算部７３０からで得られた主成分得点の正規化を行い、学習パターン（データ類型）の分類を行う。データ正規化部７４０は、データ分類部の例である。
【００２３】
収束判定部７５０は、上記データ正規化部７４０により設けられたパターン数を計数するとともに、パターンの増加状況を監視し、パターン数が収束しているか否かの判定を行う。
収束判定部７５０は、パターン数が収束するまでは、データ入力・処理部７２０の処理を継続させることで主成分得点計算部７３０に学習対象データの入力及び主成分得点の算出を継続させ、パターン数が収束したと判断した際に、データ入力・処理部７２０の処理を終了させることで主成分得点計算部７３０に学習対象データの入力及び主成分得点の算出を終了させる。
収束判定部７５０は、例えば、パターン数が収束した際にデータ入力・処理部７２０に

対して終了指示を出力してデータ入力・処理部７２０の処理を終了させることができる。また、逆に、収束判定部７５０は、例えば、パターン数が収束するまではデータ入力・処理部７２０に対して継続指示を出力して処理を継続させ、パターン数が収束した際に継続指示の出力を停止することでデータ入力・処理部７２０の処理を終了させることができる。

また、収束判定部７５０は、パターン数が収束した後に、パターンごとに、各パターンに属する主成分得点の代表となる代表値（代表データ特性値）を選択し、選択した各代表値を時系列データの異常検知に用いられる学習データとして出力する。

【００２４】

データ出力部７６０は、上記収束判定部７５０から学習データを入力し、当該学習データを学習データＤＢ８へ格納する。

【００２５】

ここで、図２１及び図２２のフローチャートを参照して、本実施の形態に係る学習データ取得部７（データ処理装置）の動作例（データ処理方法）を概説する。

なお、本実施の形態では、検査対象の時系列データの一部を学習対象データとすることとし、検査対象の時系列データが入力された際に、図２１のフローチャートに示す処理が開始し、学習データの取得が行われる。

【００２６】

先ず、データ入力・処理部７２０が、学習の対象となる時系列データである入力データ７１０を入力する（Ｓ２１０１）。前述したように、異常検知の対象となる時系列データの一部を学習対象データとして用いるため、データ入力・処理部７２０は、異常検知の対象となる時系列データの一部を入力データ７１０として入力する。

そして、データ入力・処理部７２０は、入力データ７１０を所定の単位時間ごとに集計する（Ｓ２１０２）。

その後、データ入力・処理部７２０は、集計後のデータを主成分得点計算部７３０に出力する。

【００２７】

次に、主成分得点計算部７３０が、データ入力・処理部７２０から出力されたデータを入力するとともに、入力したデータを所定の領域に区分し、領域ごとに主成分得点を算出する（Ｓ２１０３）（データ特性値算出ステップ）。

データ入力・処理部７２０からのデータは、所定の順序に従って整列されており、この順序に従ってデータを複数の領域（グループ）にグループ化し、各領域に含まれるデータのデータ値の主成分分析を行って、各グループの特徴量を算出する。

そして、主成分得点計算部７３０は、領域ごとの主成分得点を示したデータをデータ正規化部７４０に出力する。

なお、主成分得点計算部７３０で計算された主成分得点は特徴量ともいう。

【００２８】

データ正規化部７４０は、各領域の主成分得点の正規化（値をまるめる）を行った後、各領域の正規化後の主成分得点を２次元平面に配列し、主成分得点の分布から領域ごとにパターンに分類する（Ｓ２１０４）（データ分類ステップ）。

データ正規化部７４０のパターン分類の詳細は後述する。

その後、データ正規化部７４０は、各領域の主成分得点のパターン分類結果を収束判定部７５０に出力する（Ｓ２１０５）。

【００２９】

次に、データ入力・処理部７２０が、収束判定部７５０から終了指示を入力したか否かを判断し（Ｓ２１０６）、終了指示を入力していない場合は（Ｓ２１０６でＮＯ）、Ｓ２１０１に処理を戻し、時系列データを入力する。

他方、終了指示を入力した場合は（Ｓ２１０６でＹＥＳ）、時系列データの入力を終了する。

なお、ここでは、終了指示を入力することで時系列データの入力を終了することとした

が、継続指示を入力している間は時系列データの入力を継続し、継続指示の入力が終了した際に時系列データの入力を終了するようにしてもよい。

【００３０】

次に、図２２を参照して、収束判定部７５０の動作例（収束判定ステップ）を説明する。

【００３１】

収束判定部７５０は、データ正規化部７４０よりパターン分類結果を入力し（Ｓ２２０１）、パターンごとに発生回数を更新する（Ｓ２２０２）。

データ正規化部７４０からのパターン分類結果には、学習対象データの各領域のパターン（各領域の主成分得点が属するパターン）が示されている。また、収束判定部７５０は、パターンごとの発生数を管理する発生状況データテーブルを有しており、データ正規化部７４０から入力したパターン分類結果に示されている各領域のパターンの発生数を発生状況データテーブルに反映させて、発生状況データテーブルの各パターンの発生回数を更新する（増加させる）。

【００３２】

なお、パターンの発生回数又はパターンの発生数とは、あるパターンが学習対象データにおいて何回登場するかを示す（例えば、パターンX_1が学習対象データにおいて１０回登場する等）。

一方、後述するパターンの個数又はパターンの出現回数とは、学習対象データにおいていくつのパターンが含まれるかを示す（例えば、学習対象データに、パターンX_1～X_5の５個のパターンが含まれる等）。

【００３３】

次に、収束判定部７５０は、発生状況データテーブルに示されているパターンのうち発生回数が１以上のパターンの個数をカウントする（Ｓ２２０３）。

次に、収束判定部７５０は、パターンの個数が収束したか否かを判断する（Ｓ２２０４）。

収束判定部７５０は、例えば、一定時間が経過してもパターン数が増加しない場合に収束したと判断する。

また、パターン数の増加率が一定レベルまで鈍化した場合に収束したと判断してもよい。

また、想定される総パターンの大部分（例えば、９０％）が既に出現している場合に収束したと判断してもよい。

更には、これらを組み合わせてもよい。

【００３４】

収束判定部７５０は、パターン数が収束していないと判断した場合（Ｓ２２０４でＮＯ）は、Ｓ２２０１～Ｓ２２０４の動作を繰り返す。

他方、パターン数が収束したと判断した場合（Ｓ２２０４でＹＥＳ）は、収束判定部７５０は、データ入力・処理部７２０に終了指示を出力する（Ｓ２２０５）。

なお、前述したように、パターン数が収束した際に終了指示を出力する代わりに、パターン数が収束していない間は継続指示を出力し、パターン数が収束した際に継続指示の出力を停止するようにしてもよい。

【００３５】

次に、収束判定部７５０は、パターンごとに代表値を選択する（Ｓ２２０６）。

収束判定部７５０は、例えば、パターンごとに、そのパターンに属する主成分得点の平均値を算出し、平均値を代表値として選択してもよいし、パターンごとに、そのパターンに属する主成分得点の最小値、中央値、最大値のいずれかを代表値として選択してもよい。

【００３６】

次に、収束判定部７５０は、各パターンの代表値を示すデータを学習データとしてデータ出力部７６０に出力する（Ｓ２２０７）。

その後、データ出力部７６０は、学習データを学習データＤＢ８に格納し、異常検知部４が学習データＤＢ８に格納されている学習データを用いて異常検知を行う。

【００３７】

このように本実施の形態に係る学習データ取得部７では、得られた主成分得点ののべ数を集計し、その数が一定の値に収束した場合に自動的に学習を停止することで従来手法の課題を解決する。

その結果ネットワーク異常検知システムの学習動作の期間を明確化でき、ネットワーク異常検知システムの運用を自動化することが可能になる、さらに学習処理を過不足なく行うことが可能になる。

【００３８】

なお、収束判定部７５０は、図２２に示す処理に代えて図２３に示す処理を行うようにしてもよい。

つまり、図２２では、収束判定部７５０は、データ正規化部７４０からパターンの分類結果を入力する度に、パターンごとにパターンの発生回数を更新したが、これに代えて、図２３の処理では、収束判定部７５０は、データ正規化部７４０からのパターン分類結果に示されるパターンと発生状況データテーブルに示されるパターンとを比較し、パターン分類結果に発生状況データテーブルに含まれていない新規なパターンが含まれている場合（Ｓ２３０１でＹＥＳ）は、当該新規パターンを発生状況データテーブルに追加した後（Ｓ２３０２）、パターン数が収束したかどうかの判定を行う（Ｓ２２０４）。

一方、パターン分類結果に新規パターンが含まれていない場合（Ｓ２３０１でＮＯ）は、処理をＳ２２０１に戻す。

このようにしても、学習対象データにおけるパターン数の収束を検知することができる。

なお、図２３において、Ｓ２３０１及びＳ２３０２以外の処理は、図２２に示したものと同様である。

【００３９】

次に、本実施の形態に係る学習データ取得部７の動作を詳細に説明する。

【００４０】

データ入力・処理部７２０は、解析を行う対象となる入力データ７１０を単位時間ごとに集計する。初期設定のためのパラメータは以下の通りである。

集計単位時間…観測を行う時系列データを集計する単位時間

【００４１】

入力データ７１０の形式を図１０に示す。

なお、図１０に示す通し番号は各データを現すもので、説明のために記載しているものであり、実際のデータには存在しない。

入力データ７１０は、例えば送信元ＩＰアドレス毎のパケット数のデータであり、通常、このような入力データ７１０は不定期に発生するため、データ入力・処理部７２０では、あらかじめ指定した集計単位時間ごとにデータをまとめる。

図１０では、イベント発生日時（集計前イベント発生日時）は、不規則な時間間隔になっている。

【００４２】

図１１は、集計後の入力データの例である。

図１１では、イベント発生日時（集計後イベント発生日時）は単位時間に集計を開始した最初の時刻とする。また、イベント発生数（集計後イベント発生数）は単位時間に発生した集計前イベント発生数の総計である。

入力データの単位時間が、｛T_1、T_2、T_3｝、｛T_4、T_5｝、｛T_6、T_7｝に分かれる場合、集計結果は３種類の情報になる。単位時間｛T_1、T_2、T_3｝のデータを集計した結果は通し番号a_1である。集計後イベント発生日時はT_1、集計後イベント発生数はC_1からC_3を加算したものである。

なお、図１０と同様に、図１１の通し番号も説明のために付加したものであり、実際の

データには存在しない。
　また、図１１のデータは、図４に示すように、主成分得点計算部７３０に出力される。
【００４３】
　図５は、入力データ７１０を５分間隔で集計した場合の例である。
　入力データ７１０の先頭８つのイベントが集計されて５つのイベントとなる。
　入力データのうち２００７／０７／０１　０：００：２０と２００７／０７／０１　０：０１：１３、２００７／０７／０１　０：０３：０４は開始５分間に発生したイベントであるためひとつのイベントとする。
　その際イベント発生日時は先に現れた情報（２００７／０７／０１　０：００：２０）を使用し、イベント発生数は両者の合計数１７（４＋８＋５）とする。
　同様にイベント発生日時が２００７／０７／０１　０：１０：３３と２００７／０７／０１　０：１１：３０のもの、２００７／０７／０１　０：１６：２２と２００７／０７／０１　０：１９：５４のものはひとつにまとめる。
　イベントの集計時間内に１度しか発生しない場合（２００７／０７／０１　０：２２：４３）はそのまま保持し、集計時間内に１度も発生しない場合はイベント発生時間を単位時間（図１４の場合２００７／０７／０１　０：０５：００）、イベント発生数を０とする。
【００４４】
　主成分得点計算部７３０は、上記データ入力・処理部７２０で集計された時系列データから主成分得点の計算を行い、次に主成分得点の時系列へ変換する。初期設定のためのパラメータは以下の通りである。
　主成分対象次元数…主成分分析を計算する次元数
【００４５】
　主成分対象次元数は、主成分分析を計算する際の主成分対象行列の列数になる、データ入力・処理部７２０から受けた時系列データを解析する個数である。
　主成分得点計算部７３０は、時系列データの先頭から主成分対象次元数の個数のデータを取り出し主成分分析にかける。
　主成分得点計算部７３０の入力データの例を図１２に示す。
　主成分得点計算部７３０の入力データである図１２のデータと、データ入力・処理部７２０の出力データである図１１のデータは同じである。
　図１１と図１２では、以降の説明の便宜のため表記方法が異なっているが、図１１の通し番号a_1の集計後イベント発生日時T_1、集計後イベント発生数$C_1+C_2+C_3$が、図１２の通し番号d_1のイベント発生日時T_1、イベント発生数C_1に対応し、図１１の通し番号a_2の集計後イベント発生日時T_4、集計後イベント発生数C_4+C_5が、図１２の通し番号d_2のイベント発生日時T_2、イベント発生数C_2に対応する関係である。以降の行についても同様である。
【００４６】
　ここで、主成分対象次元数をｋとしたとき、時系列データの先頭からｋ個ずつまとめてグループ化し、グループごと（領域ごと）に処理を行う。図１２の例の場合d_1からd_kまでのイベント発生数から１行ｋ列の行列を作成し、この行列に含まれる要素を一つのグループ（領域）として主成分分析を行う。取り扱う行列は以下のようになる。
　　（C_1　C_2　．．．　C_k）
　その後、時系列データから次のｋ個を取り出し同様に行列を作成して主成分分析を行う。この処理を順次繰り返す。
【００４７】
　主成分分析の結果、ｋ個の時系列データを表す主成分得点の時系列が得られる。主成分得点は第１、第２、…と複数の得点が出るが、そのうち先頭２つを以降の工程で使用する。
　時系列データから作成した配列と主成分分析で得られた特徴量の関係を図１３に示す。
【００４８】

図１３において、ＰＣ$_{1_1}$およびＰＣ$_{2_1}$は、入力の時系列データから作成した配列（Ｃ$_1$　Ｃ$_2$　．．．　Ｃ$_k$）をあらわす特徴量である。以下の配列についても同様である。

【００４９】

図６は、主成分得点計算部７３０による上記の手順を時系列データで表した例である。

はじめに時系列データ（データ入力・処理部７２０による集計後の時系列データ）を先頭からｋ要素ずつ分割したｎ個の部分時系列（領域）を作成する。

次に、それぞれの部分時系列に対して主成分分析を行う。

主成分分析の概念を図７に示す。

この結果一つの部分時系列あたり２つの主成分得点が得られた。

本工程の出力として、主成分得点計算部７３０は、イベントの発生時間と特徴量を記述した図１４に示すデータを作成し、データ正規化部７４０に出力する。

【００５０】

データ正規化部７４０は、図１５に示すようなデータを入力し、上記主成分得点計算部７３０で得られた特徴領域の群を調査し、他の領域と比較して領域のスコア化を行う。なお、図１５では、説明の便宜のために通し番号を付与しているが、実際のデータにはなく、実際は、図１４と同じ形式のデータを入力する。

データ正規化部７４０による特徴領域の調査は、具体的には、上記主成分得点計算部７３０からの入力から第１特徴量と第２特徴量を取り出し、第１特徴量及び第２特徴量の正規化を行った後、２次元平面へ配置する。配置の方法は、例えば、第１特徴量をＹ軸の座標に配置し、第２特徴量をＸ軸の座標とする。

【００５１】

図８は、主成分得点計算部７３０からの入力データ（図１５）を正規化して２次元の特徴量空間（主成分空間）へ配置した図である。

通し番号（ａ）と（ｆ）は主成分空間における位置が近く、同じデータ類型とみなすことができ、通し番号（ａ）と（ｆ）の領域の主成分得点は同じパターンに分類される。

同様に、通し番号（ｂ）と（ｄ）は主成分空間における位置が近く、同じデータ類型とみなすことができ、通し番号（ｂ）と（ｄ）の領域の主成分得点は同じパターンに分類される。

また、通し番号（ｃ）と（ｅ）は主成分空間における位置が近く、同じデータ類型とみなすことができ、通し番号（ｃ）と（ｅ）の領域の主成分得点は同じパターンに分類される。

【００５２】

また、図８及び図１３～図１５では、主成分得点として主成分得点計算部７３０において２つの特徴量が採用される例を説明したが、図９及び図１４に示すように主成分得点として１つの特徴量が採用され、１つの特徴量に対してデータ正規化部７４０がパターンに分類するようにしてもよい。

【００５３】

図１７は、データ正規化部７４０から収束判定部７５０へ出力されるパターン分類結果を示すデータである。

図１７のデータでは、イベント発生日時（Ｔ$_1$～Ｔ$_{nk+1}$）ごとに、主成分得点のパターン（Ｘ$_1$～Ｘ$_{pp}$）が示される。図１７において、Ｘ$_1$、Ｘ$_2$、Ｘ$_3$、Ｘ$_4$等は、それぞれ異なるパターンであることを示す。

また、データ正規化部７４０は、主成分得点計算部７３０から入力した図１５又は図１６の特徴量のデータも収束判定部７５０に出力する。

【００５４】

収束判定部７５０は、図１５又は図１６に示す特徴量のデータ及び図１７に示すデータをデータ正規化部７４０から入力する。

そして、収束判定部７５０は、パターンごとの出現数を計数し、パターン数が収束しているか判定を行う。

図１８は、収束判定部７５０が管理している発生状況データテーブルの例を示す。図１８は、初期値が設定された発生状況データテーブル（パターン発生数及びパターン個数の計数前の発生状況データテーブル）の例を示している。

発生状況データテーブルは、パターン（X_1～X_{pp}）ごとに発生回数をカウントするためのテーブルである。

収束判定部７５０は、図１７に示すデータ正規化部７４０からのパターン分類結果中の各パターン発生数を計数し、計数結果を図１８の発生状況データテーブルに書き込む。

また、発生状況データテーブルにおいて発生数が１以上のパターンの個数を計数し、項目数の欄に書き込む。

図１９は、収束判定部７５０によりパターンごとの発生数及び項目数が書き込まれた後の発生状況データテーブルの例を示している。

図１９の例では、発生数が１以上のパターンは、X_1、X_2、X_4、X_6、X_7の５つであり、項目数に５が記入されている。

【００５５】

また、図２３に示したように、収束判定部７５０は、データ正規化部７４０からのパターン分類結果に新規なパターンが含まれていた場合に、当該新規なパターンを発生状況データテーブルに追加するようにしてもよい。

図２４は、このような場合に用いられる発生状況データテーブルの例を示している。

図２４の発生状況データテーブルでは、パターンごとの発生回数は管理しておらず、データ正規化部７４０のパターン分類結果に現れたパターン名のみを管理している。

そして、これまでX_1～X_4のパターンがデータ正規化部７４０のパターン分類結果に出現していた場合に、今回データ正規化部７４０から入力したパターン分類結果のデータにパターンX_5が含まれていた場合に、このパターンX_5は発生状況データテーブルに含まれていないので、新規なパターンであり、収束判定部７５０は、このパターンX_5を新たに発生状況データテーブルに追加する。

このような手順によっても、収束判定部７５０は学習対象データにおけるパターンの出現数をカウントすることができる。

【００５６】

そして、収束判定部７５０は、図１９の発生状況データテーブルの項目数の欄に記入されているパターンの個数（図２４の発生状況データテーブルの場合は、レコード数）に基づいてパターン個数が収束したか否かを判断する。

ここで、例えば、前回計数した項目数をＮＰＰ、閾値をＴＨと置く。

今回計数した項目数と前回計数した項目数ＮＰＰの差異が閾値ＴＨ以内であった場合、収束判定部７５０は、パターン数が収束したとみなし、終了指示をデータ入力・処理部７２０に出力し、次のデータ出力処理へ移る。

差異が閾値ＴＨ以上であった場合、学習が継続しているとみなしデータ入力・処理部７２０の処理を継続させる。

【００５７】

また、他の方法として、収束判定部７５０は、収束判定を行う度に、収束判定を行った時刻と項目数を記憶しておき、単位時間あたりのパターン個数の増加率を計算し、単位時間あたりのパターン個数の増加率が所定レベル以下（例えば、１％以下）に鈍化した場合に、パターン数が収束したと判定するようにしてもよい。

【００５８】

また、他の方法として、収束判定部７５０は、想定されるパターン総数（X_{pp}個）の所定割合（例えば、９０％）に相当するパターン数を基準値とし、パターン出現数が基準値に到達した場合に、パターン数が収束したと判定するようにしてもよい。

【００５９】

次に、収束判定部７５０は、各パターンでの代表値を選択する。

代表値は、データ正規化部７４０から入力した図１５又は図１６の特徴量データと図１７のパターン分類結果から選択する。

以下、図１６の特徴量データと図１７のパターン分類結果を用いて、代表値を選択する例を説明するが、図１６の特徴量データの代わりに図１５の特徴量データを用いる場合でも同様の処理となる。

先ず、収束判定部７５０は、例えば、イベント発生日時の項目に従って図１６の特徴量と図１７のパターンとを対応づける。

具体的には、収束判定部７５０は、図１６のイベント発生日時T_1のレコードに、図１７のイベント発生日時T_1のレコードに記述されているパターンX_1を追加し、イベント発生日時T_1のレコードにおいて特徴量P_1とパターンX_1とを対応づける。イベント発生日時T_{k+1}以降についても同様の処理を行う。イベント発生日時T_{nk+1}まで特徴量Ｐとパターンｘとが対応づけられた後、同一パターンが記述されているレコードを集め、同一パターンが記述されているレコードの特徴量の中からそのパターンの代表値を選択する。

代表値は、同一パターンを有するレコードの特徴量の平均値でもよいし、最大値、最小値、中央値等でもよい。

また、イベント発生日時が最も古いのレコードの特徴量でもよいし、イベント発生日時が最も新しいレコードの特徴量でもよい。

図２０は、収束判定部７５０によりパターンごとに選択された代表値（特徴量）を示すデータである。収束判定部７５０は、図２０に示すデータを学習データとしてデータ出力部７６０に出力する。

なお、図２０の学習データでは、選択された代表値のイベント発生日時の項目が付加されているが、イベント発生日時の項目は省略可能であり、代表値のみが示されるデータであってもよい。

また、イベント発生日時に代えて、またはイベント発生日時に加えて、代表値のパターンを示すようにしてもよい。

【００６０】

データ出力部７６０は、収束判定部７５０から図２０に示す学習データを入力し、学習データＤＢ８に格納する。

データ出力部７６０は、図２０に示す形式のまま学習データを学習データＤＢ８に格納してもよいし、学習データＤＢ８のデータフォーマットに沿うように加工してから学習データを格納するようにしてもよい。

【００６１】

以降は、異常検知部４が学習データＤＢ８に格納されている学習データを用いて、異常検知を行う。

異常検知の動作自体は、既存の手法と同様であるため、説明は省略する。

【００６２】

このように、本実施の形態では、データのパターンを調査し、パターンの出現数が一定の水準に達したことを自動的に判断する。

そして、本実施の形態に係る学習データの取得手法をネットワーク異常検知システムの前段階での適用を行うことにより、従来技術で調整が必要であった学習処理を簡便にし、特に内部ネットワークでの監視のような早期にシステムの検知体制が必要な箇所での異常検知に効果がある。

つまり、パターン出現数が収束した場合に学習を終了させることとしているので、異常検知に必要十分な学習データが収集された段階で学習データの収集を停止することになり、異常検知処理を早期に開始させることができるとともに、高精度な異常検知を行えるだけの十分な量の学習データを蓄積することができる。

【００６３】

以上、本実施の形態ではネットワークの不正アクセスを監視する方法および装置に関して説明を行った。特にＬＡＮなどの内部ネットワークで発生する通信トラフィックの監視を容易にするために監視するトラフィックの特徴を学習する学習期間を自動的に判定する方法を（１）学習データの取得（データ入力）、（２）データの主成分の得点計算、（３

）得られた得点の正規化と学習パターンの集計、（４）学習パターンの発生状況からパターンの収束判定で実現することを説明した。
そして、収束判定は、主成分の得点の延べ数が一定の値に収束した場合に学習を停止する方法によることを説明した。
【００６４】
また、本実施の形態では、学習データ取得部は、主に以下の手段を備えることを説明した。
時系列データを単位時間ごとに集計した数を記憶するデータ入力・処理部、
上記データ入力・処理部で集計された時系列データから主成分得点の時系列を計算する主成分得点計算部、
上記主成分得点計算部で得られた主成分得点の時系列の正規化を行い学習パターンの集計を行うデータ正規化部、
上記データ正規化部から得られた学習パターンの発生状況を調査し、学習パターン数が収束しているか判定を行う収束判定部、
上記収束判定部から学習したデータをデータベースへ格納するデータ出力部。
【００６５】
最後に、実施の形態１に示した不正アクセス分析システム１００及び学習データ取得部７のハードウェア構成例について説明する。
【００６６】
図２５は、実施の形態１に示す不正アクセス分析システム１００及び学習データ取得部７のハードウェア資源の一例を示す図である。なお、図２５の構成は、あくまでも不正アクセス分析システム１００及び学習データ取得部７のハードウェア構成の一例を示すものであり、不正アクセス分析システム１００及び学習データ取得部７のハードウェア構成は図２５に記載の構成に限らず、他の構成であってもよい。
【００６７】
図２５において、不正アクセス分析システム１００及び学習データ取得部７は、プログラムを実行するＣＰＵ９１１（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ、中央処理装置、処理装置、演算装置、マイクロプロセッサ、マイクロコンピュータ、プロセッサともいう）を備えている。ＣＰＵ９１１は、バス９１２を介して、例えば、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）９１３、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）９１４、通信ボード９１５、表示装置９０１、キーボード９０２、マウス９０３、磁気ディスク装置９２０と接続され、これらのハードウェアデバイスを制御する。更に、ＣＰＵ９１１は、ＦＤＤ９０４（Ｆｌｅｘｉｂｌｅ　Ｄｉｓｋ　Ｄｒｉｖｅ）、コンパクトディスク装置９０５（ＣＤＤ）、プリンタ装置９０６、スキャナ装置９０７と接続していてもよい。また、磁気ディスク装置９２０の代わりに、光ディスク装置、メモリカード読み書き装置などの記憶装置でもよい。
ＲＡＭ９１４は、揮発性メモリの一例である。ＲＯＭ９１３、ＦＤＤ９０４、ＣＤＤ９０５、磁気ディスク装置９２０の記憶媒体は、不揮発性メモリの一例である。これらは、記憶装置あるいは記憶部の一例である。
通信ボード９１５、キーボード９０２、スキャナ装置９０７、ＦＤＤ９０４などは、入力部、入力装置の一例である。
また、通信ボード９１５、表示装置９０１、プリンタ装置９０６などは、出力部、出力装置の一例である。
【００６８】
通信ボード９１５は、例えば、ＬＡＮ（ローカルエリアネットワーク）、インターネット、ＷＡＮ（ワイドエリアネットワーク）などに接続されていてもよい。
【００６９】
磁気ディスク装置９２０には、オペレーティングシステム９２１（ＯＳ）、ウィンドウシステム９２２、プログラム群９２３、ファイル群９２４が記憶されている。
プログラム群９２３のプログラムは、ＣＰＵ９１１がオペレーティングシステム９２１

、ウィンドウシステム９２２を利用しながら実行する。
【００７０】
　また、ＲＡＭ９１４には、ＣＰＵ９１１に実行させるオペレーティングシステム９２１のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。
　また、ＲＡＭ９１４には、ＣＰＵ９１１による処理に必要な各種データが格納される。
【００７１】
　また、ＲＯＭ９１３には、ＢＩＯＳ（Ｂａｓｉｃ　Ｉｎｐｕｔ　Ｏｕｔｐｕｔ　Ｓｙｓｔｅｍ）プログラムが格納され、磁気ディスク装置９２０にはブートプログラムが格納されている。
　不正アクセス分析システム１００及び学習データ取得部７の起動時には、ＲＯＭ９１３のＢＩＯＳプログラム及び磁気ディスク装置９２０のブートプログラムが実行され、ＢＩＯＳプログラム及びブートプログラムによりオペレーティングシステム９２１が起動される。
【００７２】
　上記プログラム群９２３には、実施の形態１の説明において「～部」として説明している機能を実行するプログラムが記憶されている。プログラムは、ＣＰＵ９１１により読み出され実行される。
【００７３】
　ファイル群９２４には、実施の形態１の説明において、「～の判断」、「～の計算」、「～の比較」、「～の評価」、「～の判定」、「～の設定」、「～の計数」、「～の更新」、「～の分類」、「～の集計」等として説明している処理の結果を示す情報やデータや信号値や変数値やパラメータが、「～ファイル」や「～データベース」の各項目として記憶されている。「～ファイル」や「～データベース」は、ディスクやメモリなどの記録媒体に記憶される。ディスクやメモリになどの記憶媒体に記憶された情報やデータや信号値や変数値やパラメータは、読み書き回路を介してＣＰＵ９１１によりメインメモリやキャッシュメモリに読み出され、抽出・検索・参照・比較・演算・計算・処理・編集・出力・印刷・表示などのＣＰＵの動作に用いられる。抽出・検索・参照・比較・演算・計算・処理・編集・出力・印刷・表示のＣＰＵの動作の間、情報やデータや信号値や変数値やパラメータは、メインメモリ、レジスタ、キャッシュメモリ、バッファメモリ等に一時的に記憶される。
　また、実施の形態１で説明するフローチャートの矢印の部分は主としてデータや信号の入出力を示し、データや信号値は、ＲＡＭ９１４のメモリ、ＦＤＤ９０４のフレキシブルディスク、ＣＤＤ９０５のコンパクトディスク、磁気ディスク装置９２０の磁気ディスク、その他光ディスク、ミニディスク、ＤＶＤ等の記録媒体に記録される。また、データや信号は、バス９１２や信号線やケーブルその他の伝送媒体によりオンライン伝送される。
【００７４】
　また、実施の形態１の説明において「～部」として説明しているものは、「～回路」、「～装置」、「～機器」、であってもよく、また、「～ステップ」、「～手順」、「～処理」であってもよい。すなわち、「～部」として説明しているものは、ＲＯＭ９１３に記憶されたファームウェアで実現されていても構わない。或いは、ソフトウェアのみ、或いは、素子・デバイス・基板・配線などのハードウェアのみ、或いは、ソフトウェアとハードウェアとの組み合わせ、さらには、ファームウェアとの組み合わせで実施されても構わない。ファームウェアとソフトウェアは、プログラムとして、磁気ディスク、フレキシブルディスク、光ディスク、コンパクトディスク、ミニディスク、ＤＶＤ等の記録媒体に記憶される。プログラムはＣＰＵ９１１により読み出され、ＣＰＵ９１１により実行される。すなわち、プログラムは、実施の形態１の「～部」としてコンピュータを機能させるものである。あるいは、実施の形態１殻の「～部」の手順や方法をコンピュータに実行させるものである。
【００７５】
　このように、実施の形態１に示す不正アクセス分析システム１００及び学習データ取得

部７は、処理装置たるＣＰＵ、記憶装置たるメモリ、磁気ディスク等、入力装置たるキーボード、マウス、通信ボード等、出力装置たる表示装置、通信ボード等を備えるコンピュータであり、上記したように「～部」として示された機能をこれら処理装置、記憶装置、入力装置、出力装置を用いて実現するものである。

【図面の簡単な説明】

【００７６】

【図１】実施の形態１に係る不正アクセス分析システムの構成例を示す図。

【図２】実施の形態１に係る不正アクセス分析システムと監視対象との関係を示す図。

【図３】実施の形態１に係る学習期間と従来の学習期間を説明する図。

【図４】実施の形態１に係る学習データ取得部の構成例を示す図。

【図５】実施の形態１に係るデータ入力・処理部のデータ集計処理の具体例を示す図。

【図６】実施の形態１に係る主成分得点計算部の領域化処理の具体例を示す図。

【図７】実施の形態１に係る主成分得点計算部の主成分分析処理の具体例を示す図。

【図８】実施の形態１に係るデータ正規化部の主成分空間への配置処理の具体例を示す図。

【図９】実施の形態１に係る主成分得点計算部の主成分分析処理の具体例を示す図。

【図１０】実施の形態１に係るデータ入力・処理部の集計前の入力データの例を示す図。

【図１１】実施の形態１に係るデータ入力・処理部の集計後の入力データの例を示す図。

【図１２】実施の形態１に係る主成分得点計算部の入力データの例を示す図。

【図１３】実施の形態１に係る主成分得点計算部における時系列データと特徴量の関係の例を示す図。

【図１４】実施の形態１に主成分得点計算部の出力データの例を示す図。

【図１５】実施の形態１に係るデータ正規化部の入力データの例を示す図。

【図１６】実施の形態１に係る主成分得点計算部の出力データの例を示す図。

【図１７】実施の形態１に係るデータ正規化部の出力データの例を示す図。

【図１８】実施の形態１に係る収束判定部の発生状況データテーブルの例を示す図。

【図１９】実施の形態１に係る収束判定部の発生状況データテーブルの例を示す図。

【図２０】実施の形態１に係る収束判定部の出力データの例を示す図。

【図２１】実施の形態１に係る学習データ取得部の動作例を示すフローチャート図。

【図２２】実施の形態１に係る収束判定部の動作例を示すフローチャート図。

【図２３】実施の形態１に係る収束判定部の動作例を示すフローチャート図。

【図２４】実施の形態１に係る収束判定部の発生状況データテーブルの例を示す図。

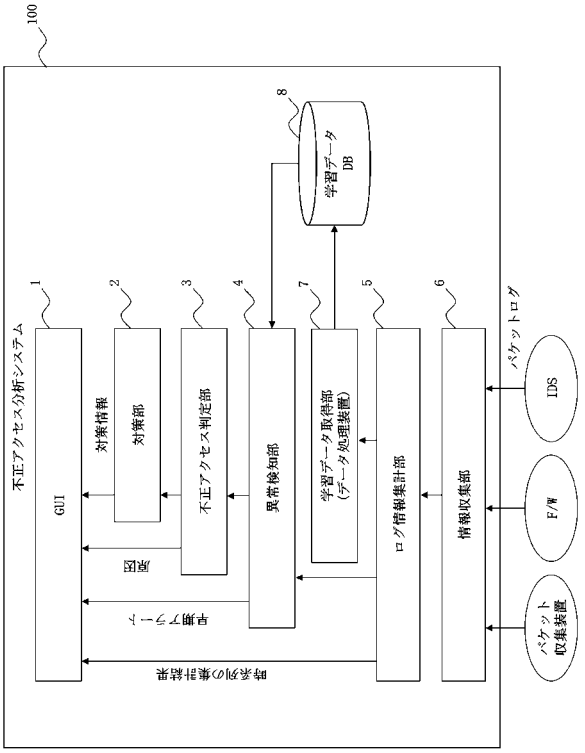
【図２５】実施の形態１に係る不正アクセス分析システム及び学習データ取得部のハードウェア構成例を示す図。

【符号の説明】

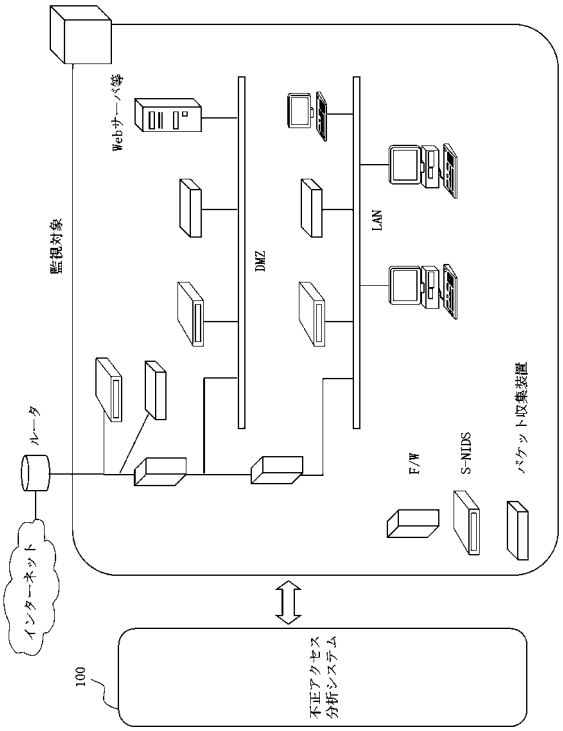
【００７７】

１　ＧＵＩ、２　対策部、３　不正アクセス判定部、４　異常検知部、５　ログ情報集計部、６　情報収集部、７　学習データ取得部、８　学習データＤＢ、１００　不正アクセス分析システム、７１０　入力データ、７２０　データ入力・処理部、７３０　主成分得点計算部、７４０　データ正規化部、７５０　収束判定部、７６０　データ出力部。

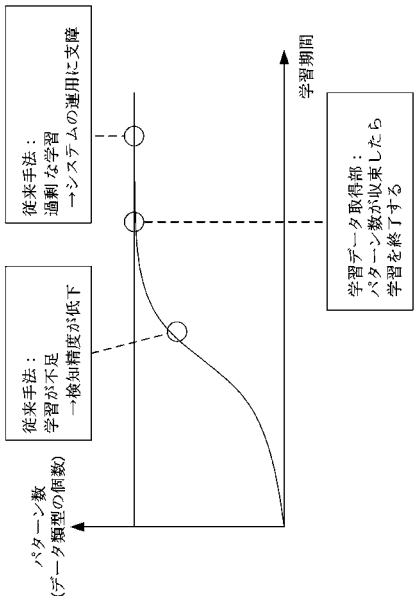
【図１】


100
不正アクセス分析システム
GUI
対策情報
対策部
不正アクセス判定部
異常検知部
学習データ取得部
(データ処理装置)
ログ情報集計部
情報収集部
学習データ
DB
原因
早期アラート
時系列の集計結果
パケットログ
IDS
F/W
パケット
収集装置
1
2
3
4
7
5
6
8

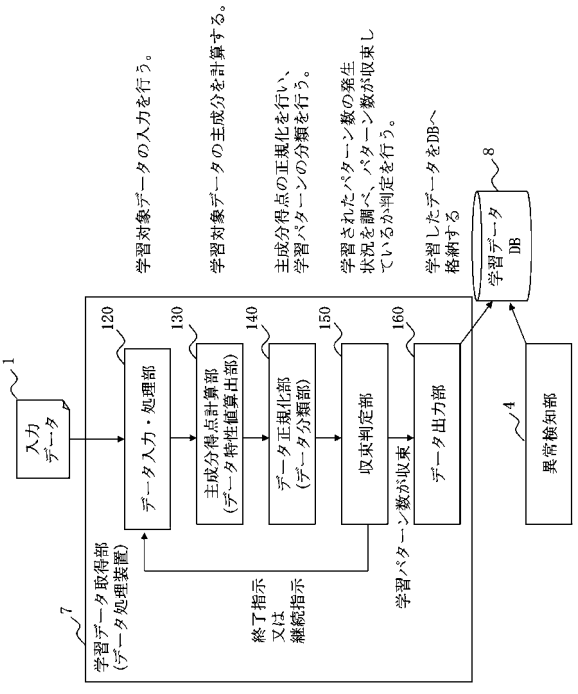
【図２】


監視対象
Webサーバ等
DMZ
LAN
ルータ
F/W
S-NIDS
パケット収集装置
インターネット
100
不正アクセス
分析システム

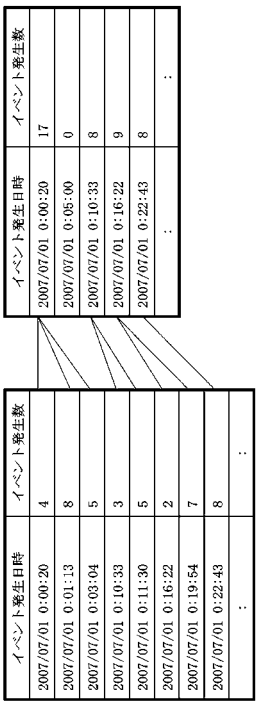
【図３】


パターン数
(データ類型の個数)
学習期間
従来手法：
学習が不足
→検知精度が低下
従来手法：
過剰な学習
→システムの運用に支障
学習データ取得部：
パターン数が収束したら
学習を終了する

【図４】


学習対象データの入力を行う。
学習対象データの主成分を計算する。
主成分得点の正規化を行い、
学習パターンの分類を行う。
学習されたパターン数の発生
状況を調べ、パターン数が収束し
ているか判定を行う。
学習したデータをDBへ
格納する
入力
データ
データ入力・処理部
主成分得点計算部
(データ特性値算出部)
データ正規化部
(データ分類部)
収束判定部
データ出力部
学習データ
DB
異常検知部
学習データ取得部
(データ処理装置)
終了指示
又は
継続指示
学習パターン数が収束
1
120
130
140
150
160
8
4
7

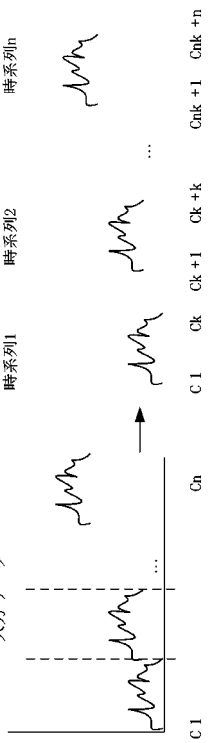
【図５】



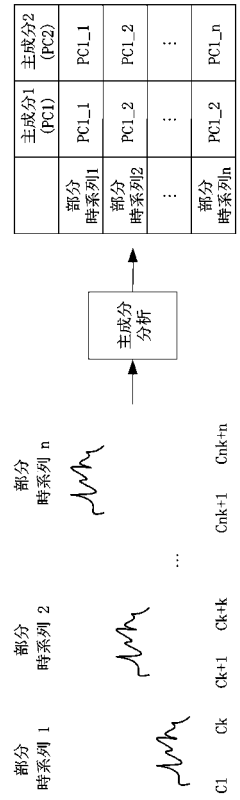
イベント発生日時	イベント発生数
2007/07/01 0:00:20	4
2007/07/01 0:01:13	8
2007/07/01 0:03:04	5
2007/07/01 0:10:33	3
2007/07/01 0:11:30	5
2007/07/01 0:16:22	2
2007/07/01 0:19:54	7
2007/07/01 0:22:43	8
：	：

イベント発生日時	イベント発生数
2007/07/01 0:00:20	17
2007/07/01 0:05:00	0
2007/07/01 0:10:33	8
2007/07/01 0:16:22	9
2007/07/01 0:22:43	8
：	：

【図６】

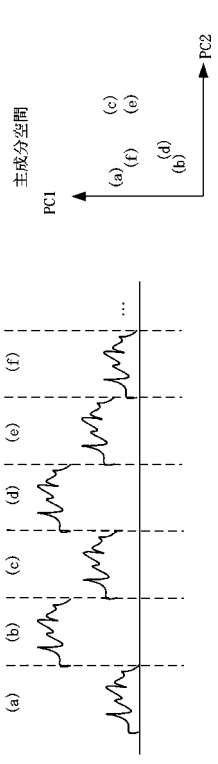


【図７】



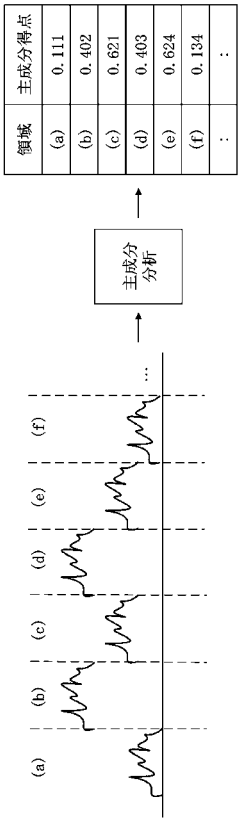
	主成分1 (PC1)	主成分2 (PC2)
部分時系列1	PC1_1	PC1_1
部分時系列2	PC1_2	PC1_2
：	：	：
部分時系列n	PC1_2	PC1_n

【図８】



領域	主成分1 (PC1)	主成分2 (PC2)
(a)	0.752	0.142
(b)	0.284	0.205
(c)	0.883	0.729
(d)	0.332	0.294
(e)	0.610	0.743
(f)	0.627	0.225
：	：	：

【図９】



領域	主成分得点
(a)	0.111
(b)	0.402
(c)	0.621
(d)	0.403
(e)	0.624
(f)	0.134
：	：

【図１０】

データ入力・処理部 集計前の入力データ

通し番号	集計前イベント発生日時	集計前イベント発生数
b_1	T_1	C_1
b_2	T_2	C_2
b_3	T_3	C_3
b_4	T_4	C_4
b_5	T_5	C_5
b_6	T_6	C_6
b_7	T_7	C_7
：	：	：

【図１１】

データ入力・処理部 集計後の入力データ

通し番号	集計後イベント発生日時	集計後イベント発生数
a_1	T_1	$C_1+C_2+C_3$
a_2	T_4	C_4+C_5
a_3	T_6	C_6+C_7
：	：	：

【図１２】

主成分得点計算部 入力データ

通し番号	イベント発生日時	イベント発生数
d_1	T_1	C_1
d_2	T_2	C_2
d_3	T_3	C_3
：	：	：
d_{k-1}	T_{k-1}	C_{k-1}
d_k	T_k	C_k
d_{k+1}	T_{k+1}	C_{k+1}
：	：	：
d_n	T_n	C_n

【図１３】

主成分得点計算部 時系列データと特徴量の関係

時系列データから作成した配列例	第1特徴量	第2特徴量
$(C_1\ C_2 \cdots C_k)$	PC_{1_1}	PC_{2_1}
$(C_{k+1}\ C_{k+2} \cdots C_{k+k})$	PC_{1_2}	PC_{2_2}
：	：	：
$(C_{nk+1}\ C_{nk+2} \cdots C_{nk+k})$	PC_{1_n}	PC_{2_n}

【図１４】

主成分得点計算部 出力データ

イベント発生日時	第1特徴量	第2特徴量
T_1	PC_{1_1}	PC_{2_1}
T_{k+1}	PC_{1_2}	PC_{2_3}
：	：	：
T_{nk-1}	PC_{1_n}	PC_{2_n}

【図１５】

データ正規化部 入力データ（収束判定部 入力データ）

通し番号	イベント発生日時	第1特徴量	第2特徴量
(a)	T_1	PC_{1_1}	PC_{2_1}
(b)	T_{k+1}	PC_{1_2}	PC_{2_2}
(c)	T_{2k+1}	PC_{1_3}	PC_{2_3}
(d)	T_{3k+1}	PC_{1_4}	PC_{2_4}
(e)	T_{4k+1}	PC_{1_5}	PC_{2_5}
(f)	T_{5k+1}	PC_{1_6}	PC_{2_6}
：	：	：	：
(n)	T_{nk+1}	PC_{1_n}	PC_{2_n}

【図１６】

主成分得点計算部 出力データ

イベント発生日時	特徴量
T_1	P_1
T_{k+1}	P_2
T_{2k-1}	P_3
T_{3k-1}	P_4
T_{4k-1}	P_5
T_{5k-1}	P_6
T_{6k-1}	P_7
：	：
T_{nk-1}	P_n

【図１７】

データ正規化部 出力データ(収束判定部 入力データ)

イベント発生日時	パターン
T_1	X_1
T_{k+1}	X_2
T_{2k-1}	X_3
T_{3k-1}	X_5
T_{4k-1}	X_2
T_{5k-1}	X_4
：	：
T_{nk-1}	X_3

【図１８】

収束判定部 発生状況データ初期値

パターン	発生数
X_1	0
X_2	0
X_3	0
X_4	0
X_5	0
X_6	0
X_7	0
：	：
X_{xy}	0
項目数(発生数が1以上の個数)	0

【図１９】

収束判定部 発生状況データ計算値

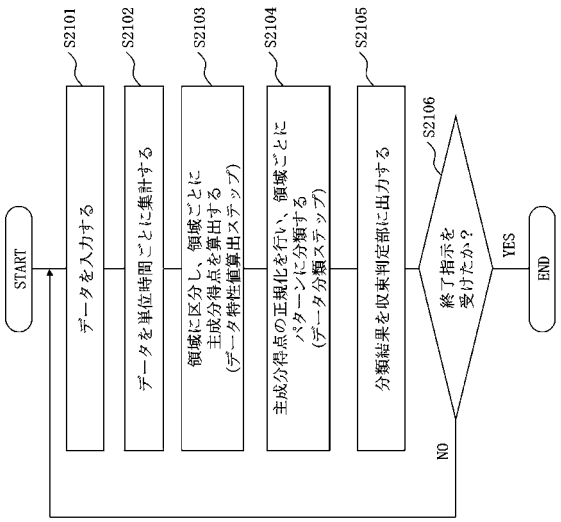
パターン	発生数
X_1	10
X_2	15
X_3	0
X_4	45
X_5	0
X_6	4
X_7	6
：	：
X_{xy}	0
項目数(発生数が1以上の個数)	5

【図２０】

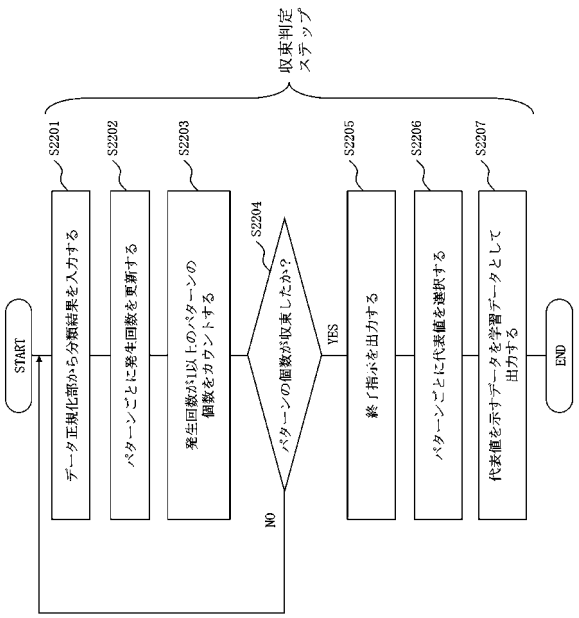
収束判定部 出力データ(データ出力部 出力データ)

イベント発生日時	特徴量
T_1	P_1
T_{k-1}	P_2
T_{2k-1}	P_1
T_{5k-1}	P_6
：	：

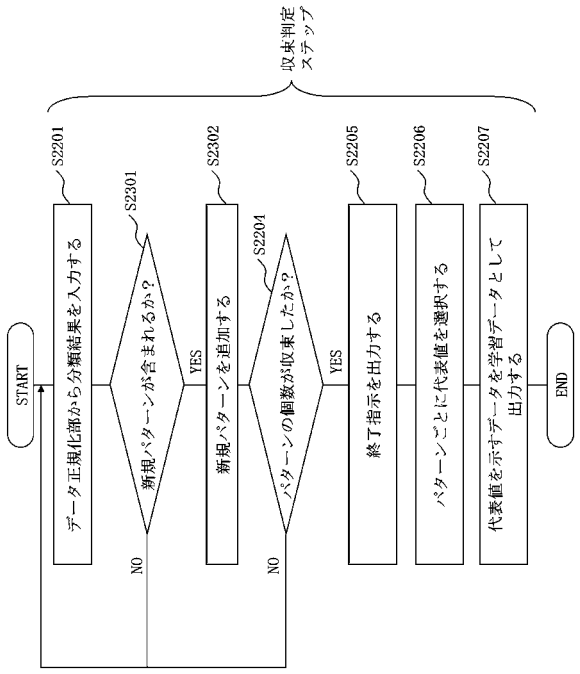
【図２１】



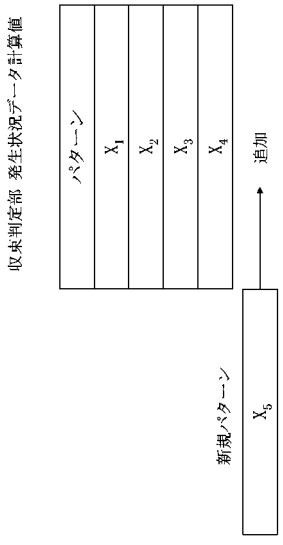
【図２２】



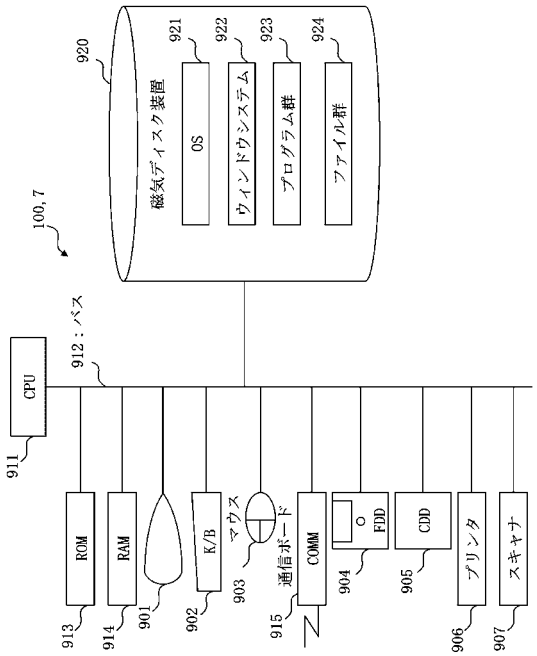
【図２３】



【図２４】



【図２５】


100,7
920
磁気ディスク装置
921
OS
922
ウィンドウシステム
923
プログラム群
924
ファイル群
912：バス
911
CPU
913
ROM
914
RAM
901
902
K/B
903
マウス
915
通信ボード
COMM
904
FDD
905
CDD
906
プリンタ
907
スキャナ

フロントページの続き

(56)参考文献 特開２００５－０８５１５７（ＪＰ，Ａ）
特開２００６－１３３９９２（ＪＰ，Ａ）
特開２００４－３１２０８３（ＪＰ，Ａ）
特開２００７－１０９１６４（ＪＰ，Ａ）
国際公開第００／５５８１１（WO，Ａ１）
特開２００７－３００２６３（ＪＰ，Ａ）
特開２００７－２３５８７９（ＪＰ，Ａ）
榊原　裕之　Hiroyuki　Sakakibara，定点観測による不正アクセス分析システム　An Intrusion Detection System by fixed-point observation of network security data，情報処理学会研究報告　Ｖｏｌ．２００６　Ｎｏ．１２９　IPSJ SIG Technical Reports，日本，社団法人情報処理学会　Information Processing Society of Japan，第2006巻

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｌ　１２／６６
Ｇ０６Ｎ　３／００
Ｈ０４Ｌ　１２／５６